Introducing Groups to an Annotation System

Amjad Hawash
hawash@di.uniroma1.it

Supervised by:
Prof. Paolo Bottoni
bottoni@di.uniroma1.it

School for advanced sciences of Luchon
Network analysis and applications
Session I, June 21 - July 5, 2014
Contents

• Web Annotation.
• MADCOW Project
• Annotations Submission: Problem & Solution
• Groups Join: Problem & Solution
• Groups-Users Matching
  – Ontology-Based Matching
  – URL-Based Matching
• Experimental Tests
• Future Work
• References.
Web Annotation: What is it?

• Associating informative data (annotations) with web resources.
• Annotations could be: text or links to multimedia documents (attachments).
• Web resources could be: text, image or video.
MADCOW Project: Architecture and services

- Multimedia Annotation of Digital Content Over the Web.

(http://www.web-annotations.com)
Annotations Submission: Problem & Solution

- Annotations (private/public).
- Problem: Privacy-Collaboration Conflict.
- Solution: Introducing Groups (with services: join types, isolation, search, operations).

Groups Join: Problem & Solution

• Problem: Manual Groups Join (Time, Effort, Irrelevance).

• Solution: Groups-Users Matching
  – Ontology-based:
    • Class Match Measure: amount of ontology coverage for a term.
    • Degree Centrality (Social Networks Analysis): quantifies the importance of a concept in an ontology with respect to its number of connections.
  – URL-Matching.
Ontology-Based Matching:
Groups-Domain-Ontology Association

- Domain-Ontology.
- Domain-Group.
Ontology-Based Matching: Class Match & Degree Centrality Measures

- Group-Domains Suggestions.
- Group-Users Suggestions.
- User-Groups Suggestions.

URL-Based Matching

• Matching the URLs annotated by both group members and non-group users.

Set of URLs annotated by the user
Experimental Tests: Introducing Groups (Collaboration, Groups' Services & Operations)

• Increased Collaboration (public 3.2, Group 5.3).
• Emerge of Invitation Time & Effort Problems.

<table>
<thead>
<tr>
<th></th>
<th>Create</th>
<th>Update</th>
<th>Invite</th>
<th>Join</th>
</tr>
</thead>
<tbody>
<tr>
<td># of times</td>
<td>72</td>
<td>51</td>
<td>719</td>
<td>125</td>
</tr>
<tr>
<td>Average (sec.)</td>
<td>37.3</td>
<td>15.9</td>
<td>99.25</td>
<td>5.6</td>
</tr>
</tbody>
</table>

Experimental Tests: Time Reduction

• Ontology Repository: 6 different Ontologies (Animals, Plants, viruses, AI, Finance, Vehicles).

• Average invitation duration is decreased from 99.25 to 10.6 seconds.

Experimental Tests: Enhanced Matching Results

- Creating dedicated ontologies (graphs) from BabelNet (http://www.babelnet.org).
- DC is preferred to CMM.
### Experimental Tests: Enhanced Matching Results

<table>
<thead>
<tr>
<th></th>
<th>Adequate</th>
<th>Presence</th>
<th>Absence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CMM</strong></td>
<td>G. Asso.</td>
<td>66%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>M. Suggestion</td>
<td>76%</td>
<td>22%</td>
</tr>
<tr>
<td><strong>DC</strong></td>
<td>G. Asso.</td>
<td>82%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>M. Suggestion</td>
<td>92%</td>
<td>3%</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td>Needed (M. Sugg.)</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>Not needed (M. Sugg.)</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

#### Table 1. Owners’ assessment of measures.

<table>
<thead>
<tr>
<th></th>
<th>Adequate</th>
<th>Presence</th>
<th>Absence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CMM G. Suggestion</strong></td>
<td>56%</td>
<td>16%</td>
<td>28%</td>
</tr>
<tr>
<td><strong>DC G. Suggestion</strong></td>
<td>86%</td>
<td>9%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>URL G. Suggestion</strong></td>
<td>Needed</td>
<td>72%</td>
<td>28%</td>
</tr>
<tr>
<td></td>
<td>Not needed</td>
<td>28%</td>
<td>28%</td>
</tr>
</tbody>
</table>
Experimental Tests: Enhanced Matching Results

Owners Ontological Measures Perceptions

- CMM G. Asso.
- User Suggestion
- DC G. Asso.
- User Suggestion

Users Ontological Measures Perceptions

- CMM G. Suggestion
- DC G. Suggestion
## Experimental Tests: Enhanced Matching Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CMM</strong></td>
<td>52%, 24%</td>
<td>20%, 30%</td>
<td>45%, 15%</td>
</tr>
<tr>
<td><strong>DC</strong></td>
<td>75%, 25%</td>
<td>66%, 16%</td>
<td>75%, 14%</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td>80%, 10%</td>
<td>50%, 16%</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. 1st and 2nd selections.

Future Works

• Studying better matching threshold.
• Try other matching measurements like: Term Frequency–Inverse Document Frequency.
• Try Harmonic Distance.
• Multiple Domain Association.
• Enhancing Groups and Users Ranking by Fuzzy Logic (why?).
References

References